Transformations of multivariate distributions

We will concentrate here on the bivariate case, but the theory described extends to the more general case.

The framework

We are interested in the situation where:

- we have two jointly distributed continuous random variables, say X and Y with joint p.d.f. $f_{X,Y}(x, y)$
- we transform them into two new continuous random variables, say U and V with joint p.d.f. $f_{U,V}(u, v)$, given by say $U = g(X, Y)$ and $V = h(X, Y)$
- the whole transformation is continuous, differentiable and **one-to-one** in that there exist "inverse" functions G and H with $X = G(U, V)$ and $Y = H(U, V)$.

The Jacobian

If we take a small region around (x, y) then this is transformed into a small region around (u, v), where $u = g(x, y)$ and $v = h(x, y)$, and the area of this new region will be the area of the old region multiplied by the **Jacobian** of the transformation

$$\left| \det \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \right|$$

evaluated at (x, y).

Change of variables

Bearing in mind that a probability density function measures probability per unit area, in order to evaluate the joint p.d.f. of U and V at (u, v) we need to take the joint p.d.f. of X and Y at (x, y) and **divide** it by this Jacobian.

Change of variables

In fact, since the joint p.d.f. is to be expressed in terms of u and v, it is (in most cases) easier to multiply by the Jacobian of the inverse transformation $x = G(u, v), y = H(u, v)$ where G and H are as defined above.

So we get

$$f_{U,V}(u, v) = f_{X,Y}(G(u, v), H(u, v)) \left| \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \right| .$$

This formula generalises the one obtained in the univariate case for monotonic g in Chapter 1.

Ranges for U and V

It is important to identify the range of values taken by (U, V), possibly with the aid of a graph.

In particular, if X and Y take values in a restricted range given by inequalities in x and y, then these must be translated into inequalities in u and v by substituting for x and y in terms of u and v.

Examples

Example 17: Transforming bivariate random variables

Example 18: Application to simulation of normal random variables

Only one new variable

Sometimes we are interested in only one transformed random variable, $U = g(X, Y)$ say.

In this case one possibility is

- to choose V arbitrarily (but not identical to or functionally dependent on U, to ensure that the joint distribution of U and V is genuinely two-dimensional),
- to find the joint p.d.f. of U and V,
- to eliminate the unwanted V by finding the marginal p.d.f. of U.

If there is no other obvious choice, choosing $V = X$ or $V = Y$ often works well.

Example

Example 19: Finding the distribution of a sum of Gamma random variables

Relationship to integration

Note that the method introduced in this section is closely related to the method used when changing variables in multiple integration.

The Student t distribution

The **Student t distribution** arises when we have independent random variables $Z \sim N(0,1)$ and $W \sim \chi^2_n$, and we consider the random variable

$$X = \frac{Z}{\sqrt{\frac{W}{n}}}.$$

We write $X \sim t_n$.

As with the chi squared distribution, the parameter n is referred to as the number of degrees of freedom.

Some properties of the t distribution

The probability density function of X can be derived using a bivariate transformation (in notes) and is

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

If $n \geq 2$ the mean $E(X) = 0$ and if $n \geq 3$ the variance $\mathrm{Var}(X) = \frac{n}{n-2}$.

As $n \to \infty$ $f(x)$ converges to the p.d.f. of a standard normal distribution.

The t distribution and the t test

You will have seen this distribution before, in MAS113 (sections 6 and 7), where the t test was introduced.

The reason it arises there is that it can be shown (see exercise 37) that if $X_1, X_2, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ random variables, the sample mean $\bar{X} \sim N(\mu, \sigma^2/n)$ and the sample variance S^2 satisfies $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, and that $\bar{X}$ and S^2 are independent.

Hence the t statistic

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}} \sim t_{n-1}.$$

The Cauchy distribution

The special case of the t distribution where $n = 1$ is the **Cauchy distribution**, seen earlier in the course.

Covariance matrices

Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)^T$ be a random (column) vector with **mean vector**

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)^T = (E(X_1), E(X_2), \ldots, E(X_k))^T = E(\boldsymbol{X}).$$

Then the $k \times k$ matrix Σ with elements given by

$$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) = E((X_i - \mu_i)(X_j - \mu_j))$$

for $i, j = 1, 2, \ldots, k$ is called the **covariance matrix** of $\mathbf{X}$, denoted by $\mathrm{Cov}(\mathbf{X})$.

Entries of the covariance matrix

- This matrix has the variances $\sigma_1^2, \sigma_2^2, \dots \sigma_k^2$ of the random variables down the diagonal
- The matrix is **symmetric**, because $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$.
- From the definition of correlation coefficient we may also write $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ where ρ_{ij} is the correlation coefficient between X_i and X_j.

More on the covariance matrix

We may also write

$$\mathrm{Cov}(\mathbf{X}) = E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T)$$

where the expectation is taken componentwise.

If $X_1, X_2, \ldots X_k$ are **independent** (or merely uncorrelated) then Σ is a **diagonal** matrix (having zero off-diagonal elements).

Example

Example 20: Example of a covariance matrix

Linear transformations of random vectors

Matrix notation is useful when we consider linear transformations of $\mathbf{X}$.

Let A be a fixed $m \times k$ matrix and $\mathbf{b}$ be a fixed m-vector, and write

$$\mathbf{Y} = A\mathbf{X} + \mathbf{b}$$

so that $\mathbf{Y}$ has m components.

Transforming the mean

Then since pre-multiplying by a matrix is a linear operation we get

$$\begin{aligned} E(\mathbf{Y}) &= AE(\mathbf{X}) + \mathbf{b} \\ E(\mathbf{Y}) &= A\boldsymbol{\mu} + \mathbf{b}. \end{aligned}$$

Transforming the covariance matrix

Also

$$\mathrm{Cov}(\mathbf{Y}) = A\,\mathrm{Cov}(\mathbf{X})A^T = A\Sigma A^T.$$

(derivation in notes)

Example

Example 21: Linear transformation of a random vector

Variance of linear combinations

Aim: find a formula for the variance of a linear combination of the random variables $X_1, X_2 \ldots, X_k$, say $Y = a_1X_1 + a_2X_2 + \ldots + a_kX_k + b$.

We do this by choosing $m = 1$ and letting A be a row vector with appropriate entries, $\mathbf{a}^T = (a_1, a_2, \ldots, a_k)$, so that $A\mathbf{X} + b$ is the scalar $Y = a_1X_1 + a_2X_2 + \ldots + a_kX_k + b$.

Using the previous theory, we get

$$\mathrm{Var}(Y) = \mathbf{a}^T\Sigma\mathbf{a}.$$

Positive definite matrices

Since this is always non-negative, we have shown that Σ is a **positive semi-definite** matrix,

(I.e. one for which $\mathbf{a}^T \Sigma \mathbf{a} \geq 0$ for all $\mathbf{a}$.)

(A **positive definite** matrix is one where the inequality is strict for all non-zero $\mathbf{a}$.)

A positive semi-definite matrix has all its eigenvalues non-negative (which can be seen by letting $\mathbf{a}$ be an eigenvector in the definition).

Variance of a sum

A particular special case is the general formula for variance of a sum $X_1 + X_2 + \ldots + X_k$, covering cases where the variables in the sum are not necessarily independent.

To do this, let each element of $\mathbf{a}$ be 1 and let $b = 0$. Then

$$\begin{aligned}
\mathrm{Var} \sum_{i=1}^{k} X_i &= (1, 1, \ldots, 1)\Sigma(1, 1, \ldots, 1)^T \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_{ij} \\
&= \sum_{i=1}^{k} \mathrm{Var}(X_i) + 2 \sum_{i,j:1 \leq i < j \leq k} \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

Example

Example 22: Variance of a sum